AD-765 176

AN INTRODUCTION TO THE MATHEMATICS OF LINEAR PREDICTIVE FILTERING AS APPLIED TO SPEECH ANALYSIS AND SYNTHESIS

Edward M. Hofstetter

Massachusetts Institute of Technology

Prepared for:

Advanced Research Projects Agency

12 July 1973

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Lincoln Laboratory, M.I.T.	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

An Introduction to the Mathematics of Linear Predictive Filtering as Applied to Speech Analysis and Synthesis

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Note

5. AUTHOR(S) (Last name, first name, initial)

Hofstetter, Edward M.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
12 July 1973	30	10

8a. CONTRACT OR GRANT NO. F19628-73-C-0002 b. PROJECT NO. ARPA Order 2006 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Note 1973-36 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) ESD-TR-73-186

10. AVAILABILITY/LIMITATION NOTICES

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
None	Advanced Research Projects Agency, Department of Defense

13. ABSTRACT

This note presents a tutorial survey of the mathematics that is used in the study of linear predictive filtering as applied to the analysis and synthesis of speech. Speech is modelled as the output of an all-pole filter that is driven by either a periodic pulse train or white noise. A minimum-mean-squared-error technique for estimating the coefficients of this filter from speech data is presented. This technique leads to a set of equations for the coefficient estimates which can be solved by a computationally efficient recursive technique known as Levinson's method.

The filter derived by the above mentioned technique can be realized by any standard technique; however, a particularly interesting realization is in terms of a digital simulation of a non-uniform acoustic tube. It is shown that any stable all-pole filter can be realized as an acoustic tube and, moreover, that the Levinson recursion produces as a by-product exactly the reflection coefficients needed for such a realization.

The report concludes by showing how the classical theory of orthogonal polynomials can be applied to the speech analysis/synthesis problem and used to derive many of the results obtained above by other means.

14. KEY WORDS

linear predictive filtering	all-pole filter
speech processing	recursive techniques

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

AN INTRODUCTION TO THE MATHEMATICS OF LINEAR PREDICTIVE FILTERING AS APPLIED TO SPEECH ANALYSIS AND SYNTHESIS

E. M. HOFSTETTER

Group 24

TECHNICAL NOTE 1973-36

12 JULY 1973

LEXINGTON MASSACHUSETTS

II

ABSTRACT

This note presents a tutorial survey of the mathematics that is used in the study of linear predictive filtering as applied to the analysis and synthesis of speech. Speech is modelled as the output of an all-pole filter that is driven by either a periodic pulse train or white noise. A minimum-mean-squared-error technique for estimating the coefficients of this filter from speech data is presented. This technique leads to a set of equations for the coefficient estimates which can be solved by a computationally efficient recursive technique known as Levinson's method.

The filter derived by the above mentioned technique can be realized by any standard technique; however, a particularly interesting realization is in terms of a digital simulation of a non-uniform acoustic tube. It is shown that any stable all-pole filter can be realized as an acoustic tube and, moreover, that the Levinson recursion produces as a by-product exactly the reflection coefficients needed for such a realization.

The report concludes by showing how the classical theory of orthogonal polynomials can be applied to the speech analysis/synthesis problem and used to derive many of the results obtained above by other means.

Accepted for the Air Force
Joseph J. Whelan, USAF
Acting Chief, Lincoln Laboratory Liaison Office

INTRODUCTION

The purpose of this note is to present a tutorial discussion of the mathematical theory underlying the analysis and synthesis of speech by means of linear predictive filtering. None of the results presented here are new, all having appeared either in the literature or in research reports. The main reason for the present note is to present these scattered results from a unified standpoint and, in some cases, to provide more detail than is available in the literature.

The basic speech problem under consideration can be formulated as follows.* Samples of a speech waveform are modelled as being the output of a digital filter that has been excited by either a series of equally spaced pulses or white noise depending whether the speech is voiced or unvoiced. The filter is described by the difference equation

$$s_n = \sum_{k=1}^{p} a_k s_{n-k} + u_n \tag{1}$$

where u_n denotes the n^{th} sample of the excitation and s_n denotes the n^{th} sample of speech. The filter order p is assumed to be known on the basis of other considerations. The transfer function of this filter is easily seen to be $\left[H_p(z)\right]^{-1}$ where

$$H_p(z) = 1 - \sum_{k=1}^{p} a_k z^{-k} \tag{2}$$

from which it is apparent that $\left[H_p(z)\right]^{-1}$ is an all-pole filter. The problem at hand is to use samples of real speech to arrive at an estimate of the filter coefficients a_k and then to use these coefficients to synthesize a filter that could be used to regenerate the original speech. The latter operation requires a knowledge of whether the original speech was voiced or unvoiced but the problem of how to obtain this information is not the concern of the present work.

There are many ways one could go about estimating the filter coefficients from the speech samples. The particular method that will be considered in this

* This section is based on references 1, 2, 6, 8, 9.

note is a minimum-mean-squared error technique that now will be described.

Select a group of N+1 speech samples which, for convenience, will be numbered from n = 0 to n = N. Define a sequence s_n by

$$s_n = \begin{cases} \text{speech sample} & 0 \le n \le N \\ 0 & n < 0,\ n > N \end{cases} \tag{3}$$

and define the mean-squared prediction error by

$$e = \sum_{n=-\infty}^{\infty} \left[s_n - \sum_{k=1}^{p} a_k s_{n-k} \right]^2 \tag{4}$$

The quantity e is a function of the assumed values for the a_k's. The desired estimate for the a_k's is obtained by choosing those values that yield a minimum value of e.

This problem can be solved by first expanding equation (4) as follows* :

$$e = \sum_n s_n^2 - 2 \sum_{k=1}^{p} a_k \sum_n s_n s_{n-k} + \sum_{k,\, j=1}^{p} a_k a_j \sum_n s_{n-k} s_{n-j}$$

$$= R_o - 2 \sum_{k=1}^{p} a_k R_k + \sum_{k,\, j=1}^{p} a_k a_j R_{k-j} \tag{5}$$

where the autocorrelation function R_k is defined by

$$R_k = R_{-k} = \sum_n s_n s_{n-k} \tag{6}$$

It will be convenient to rewrite equation (5) in matrix form as follows:

$$e = R_o - 2 \underline{a}^T \underline{r} + \underline{a}^T R\, \underline{a} \tag{7}$$

where

$$\left. \begin{aligned} \underline{a}^T &= \left[a_1, \ldots a_p \right] \\ \underline{r}^T &= \left[R_1, R_2, \ldots R_p \right] \end{aligned} \right\} \tag{8}$$

---

* All sums without limits will henceforth be assumed to run from $n = -\infty$ to $n = \infty$.

and the correlation matrix R has as its $(i,j)^{th}$ element R_{i-j}. Note that because R is a correlation matrix, it is positive definite and, therefore, non-singular.

Completing the square in equation (7) yields the result,

$$e = (\underline{a} - R^{-1}\underline{r})^T R(\underline{a} - R^{-1}\underline{r}) + R_o - \underline{r}^T R^{-1}\underline{r} \qquad (9)$$

Equation (9) may be verified simply by multiplying out the quadratic form and cancelling the appropriate terms. The desired minimization can now be performed by noting that since R is positive definite the minimum value of the quadratic form in equation (9) is zero and can be achieved by setting $\underline{a}$ equal to $\underline{a}^{(p)}$ where

$$\underline{a}^{(p)} = R^{-1}\underline{r} \qquad (10)$$

The resulting minimum e is given by

$$\begin{aligned} e_{min} \equiv e^{(p)} &= R_o - \underline{r}^T R^{-1}\underline{r} \\ &= R_o - \underline{r}^T \underline{a}^{(p)} \\ &= R_o - \sum_{k=1}^{p} a_k^{(p)} R_k \end{aligned} \qquad (11)$$

The use of the superscript p to denote the minimizing a_k's and e_{min} may seem peculiar but the reason for this notation will become apparent in the next section.

Equation (10) expresses the solution to a set of linear equations in matrix notation. In ordinary notation, the equations to which equation (10) is the solution are

$$R_i - \sum_{k=1}^{p} a_k^{(p)} R_{i-k} = 0 \qquad (12)$$

$$i = 1, \ldots, p$$

These equations, called the autocorrelation normal equations, will play a vital role in the sequel.

THE LEVINSON RECURSION

The autocorrelation normal equations (12) can be solved in a recursive way by means of a technique known as Levinson's method.* To derive this technique, first assume that the solution to the n^{th} order autocorrelation normal equations is known and denote it by $a_k^{(n)}$, $k = 1, \ldots n$. Next, write down the $n+1^{st}$ order equations in the form

$$\left.\begin{aligned} R_i - \sum_{k=1}^{n} a_k^{(n+1)} R_{i-k} - a_{n+1}^{(n+1)} R_{i-n-1} &= 0 \\ i = 1, \ldots n. & \\ R_{n+1} - \sum_{k=1}^{n} a_k^{(n+1)} R_{n+1-k} - a_{n+1}^{(n+1)} R_o &= 0 \end{aligned}\right\} \qquad (13)$$

A neat way of getting at the Levinson recursion is to assume a solution to (13) of the form

$$a_k^{(n+1)} = a_k^{(n)} - b_k \quad , \; k = 1, \ldots n. \qquad (14)$$

with $a_{n+1}^{(n+1)}$ to be determined later. Substitution of (14) into the first n of equations (13) leads to the new equation

$$\sum_{k=1}^{n} b_k R_{i-k} - a_{n+1}^{(n+1)} R_{i-n-1} = 0 \qquad (15)$$

$$i = 1, \ldots n.$$

Motivated by the fact that equations (15) look very much like the n^{th} order autocorrelation normal equations, the change of variable $j = i-n-1$ is made with the result

$$a_{n+1}^{(n+1)} R_j - \sum_{k=1}^{n} b_k R_{j+k-n-1} = 0 \qquad (16)$$

$$j = 1, \ldots n.$$

Next, the change of variable $\ell = n+1-k$ is made and (16) becomes

$$a_{n+1}^{(n+1)} R_j - \sum_{\ell=1}^{n} b_{n+1-\ell} R_{j-\ell} = 0 \qquad (17)$$

$$j = 1, \ldots n.$$

Since, equations (17) are a scaled version of the n^{th} order autocorrelation normal equations their solution is evidently given by,

* See reference 7.

$$b_{n+1-\ell} = a_{n+1}^{(n+1)} \, a_{\ell}^{(n)} \quad , \quad \ell = 1, \ldots n. \tag{18}$$

and, therefore

$$a_k^{(n+1)} = a_k^{(n)} - a_{n+1}^{(n+1)} \, a_{n+1-k}^{(n)} \tag{19}$$

It only remains to see if a value of $a_{n+1}^{(n+1)}$ can be found such that the last remaining equation in the set (13) can be satisfied. Using (19), this equation now reads,

$$R_{n+1} - \left[\sum_{k=1}^{n} a_k^{(n)} - a_{n+1}^{(n+1)} \, a_{n+1-k}^{(n)} \right] R_{n+1-k} - a_{n+1}^{(n+1)} R_o = 0 \tag{20}$$

This equation can be solved for $a_{n+1}^{(n+1)}$ with the result,

$$a_{n+1}^{(n+1)} \equiv K_n = \frac{R_{n+1} - \sum\limits_{k=1}^{n} a_k^{(n)} R_{n+1-k}}{R_o - \sum\limits_{k=1}^{n} a_k^{(n)} R_k} \tag{21}$$

This result is meaningful as long as the denominator is not zero; however, the denominator is exactly equal to the minimum mean squared error for the n^{th} stage of the process, $e^{(n)}$ as given by equation (11). However, $e^{(n)}$ can never be zero, for if it were, it would follow that $s_n = \sum\limits_{k=1}^{n} a_k s_{n-k}$ for all n. Since $s_n = 0$ for $n<0$, this equation implies that $s_n = 0$ for all n. Since this case never arises in practice, it follows that equation (21) is always meaningful.

The only ingredient missing to set this recursive process in motion is a solution to the first order system and this can be written down by inspection of (12) as

$$a_1^{(1)} \equiv K_o = \frac{R_1}{R_o} \tag{22}$$

For later considerations, it will be useful to rewrite the Levinson recursion in terms of the inverse filter transfer function $H_n(z)$ instead of in terms of the coefficients $a_k^{(n)}$ as given by equations (19) and (21). This recursion is easily seen to be given by,

$$H_{n+1}(z) = H_n(z) - K_n z^{-(n+1)} H_n(z^{-1}) \tag{23}$$

with K_n being determined by the R_k's via equation (21). The initial condition for (23) is given by

$$H_1(z) = 1 - \frac{R_1}{R_0} z^{-1} \tag{24}$$

It is evident from equation (22) that $|K_0| < 1$ and it turns out that this is true for K_n for all n. Since this fact will be vital in the sequel it will be proved now.

To this end, it will be necessary to rewrite equation (21) in the z-transform domain by making use of the easily verified identity.

$$\begin{aligned} R_k &= \sum_n s_n s_{n-k} \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-j2\pi kf} \left| S(e^{j2\pi f}) \right|^2 df \end{aligned} \tag{25}$$

where S(z) denotes the z-transform of the speech samples

$$S(z) = \sum_n s_n z^{-n} \tag{26}$$

In order to simplify notation, equation (25) will be rewritten as

$$R_k = \int z^{-k} \left| S(z) \right|^2 df \tag{27}$$

where the convention in force here and in the sequel is that all integrals have limits $(-\frac{1}{2}, \frac{1}{2})$ and whenever the variable z appears under an integral sign, it is understood to be equal to $e^{j2\pi f}$. Equation (21) which defines K_n now can be rewritten in the form

$$K_n = \frac{\int |S(z)|^2 \left[z^{-(n+1)} - \sum_{k=1}^{n} a_k^{(n)} z^{-(n+1-k)} \right] df}{\int |S(z)|^2 \left[1 - \sum_{k=1}^{n} a_k^{(n)} z^{-k} \right] df}$$

$$= \frac{\int |S(z)|^2 z^{-(n+1)} H_n(z^{-1})\, df}{\int |S(z)|^2 H_n(z)\, df} \tag{28}$$

Since the denominator of this equation is the minimum mean squared error, it follows that,

$$e^{(n)} = \int |S(z)|^2 H_n(z) \quad df \tag{29}$$

A recursion for $e^{(n)}$ can easily be derived by writing

$$\begin{aligned} e^{(n+1)} &= \int |S(z)|^2 H_{n+1}(z)\, df \\ &= \int |S(z)|^2 \left[H_n(z) - K_n z^{-(n+1)} H_n(z^{-1}) \right] df \\ &= e^{(n)} - K_n \int S(z)^2 z^{-(n+1)} H_n(z^{-1}) \quad df \\ &= e^{(n)} \left[1 - K_n^2 \right] \end{aligned} \tag{30}$$

where the last step follows from equation (28).

Since $e^{(n)}$ must always be positive, it follows from the last equation that $|K_n| < 1$ as advertised.

As an important application of the result that $|K_n| < 1$, it will be shown that all the zeros of $H_n(z)$ lie strictly inside the unit circle, which implies that the speech synthesis filters $\left[H_n(z)\right]^{-1}$ will always be stable. The proof proceeds by induction by first noting that because a correlation function is always maximum at the origin, $|R_k| < R_o$, it follows that $H_1(z)$ as defined by equation (24), has its zero inside the unit circle. Next, assume that $H_n(z)$ has its n zeros inside the unit circle. Multiplying equation (23) by z^{n+1} and noting that, on the unit circle $|z^{n+1}H_n(z)| = |H_n(z^{-1})|$, it follows from Rouche's* theorem that $z^{n+1}H_{n+1}(z)$ and $z^{n+1}H_n(z)$ have equal numbers of zeros inside the unit circle. Since $z^{n+1}H_n(z)$ has n+1 zeros inside the unit circle the proof of the statement follows by induction.

---

* Reference 10, p. 116.

The Nonuniform Acoustic Tube

Figure 1 depicts three sections of a nonuniform acoustic tube.* The cross-sectional area of the n^{th} section is A_n and the length of all sections is Δ. The forward and backward components of the volume velocity measured at the left-hand end of the n^{th} section are sampled every $2\Delta/c$ seconds and the z-transforms of these samples are denoted by $V_n^+(z)$ and $V_n^-(z)$. The constant c denotes the velocity of sound in the tube.

The relationship between the volume velocities in the n^{th} and $n+1^{th}$ sections can be determined by writing down the continuity equations for volume velocity and acoustic pressure at the boundary between the n^{th} and $n+1^{th}$ sections. The z-transforms of the forward and backward volume velocities measured at the right-hand end of the n^{th} section are given by $z^{-\frac{1}{2}}V_n^+(z)$ and $z^{\frac{1}{2}}V_n^-(z)$ respectively. The continuity of volume velocity can now be expressed by the equation

$$V_{n+1}^+(z) - V_{n+1}^-(z) = z^{-\frac{1}{2}}V_n^+(z) - z^{\frac{1}{2}}V_n^-(z) \tag{31}$$

Since the acoustic impedence of the n^{th} section is given by $\rho c/A_n$ where ρ denotes the density of air, the continuity of acoustic pressure is expressed by the equation,

$$\frac{\rho c}{A_{n+1}}\left[V_{n+1}^+(z) + V_{n+1}^-(z)\right] = \frac{\rho c}{A_n}\left[z^{-\frac{1}{2}}V_n^+(z) + z^{\frac{1}{2}}V_n^-(z)\right] \tag{32}$$

These equations can be solved for $V_{n+1}^+(z)$ and $V_{n+1}^-(z)$ with the result,

$$\begin{aligned} V_{n+1}^+(z) &= \frac{1}{1+r_n}\left[z^{-\frac{1}{2}}V_n^+(z) - r_n z^{\frac{1}{2}}V_n^-(z)\right] \\ V_{n+1}^-(z) &= \frac{1}{1+r_n}\left[-r_n z^{-\frac{1}{2}}V_n^+(z) + z^{\frac{1}{2}}V_n^-(z)\right] \end{aligned} \tag{33}$$

where the reflection coefficient r_n is defined by

$$r_n = \frac{A_n - A_{n+1}}{A_n + A_{n+1}} \tag{34}$$

---

* This section is based primarily on reference 1. Note carefully that the numbering of the tube sections differs from that in ref. 1 in that n here corresponds to Wakita's N-n.

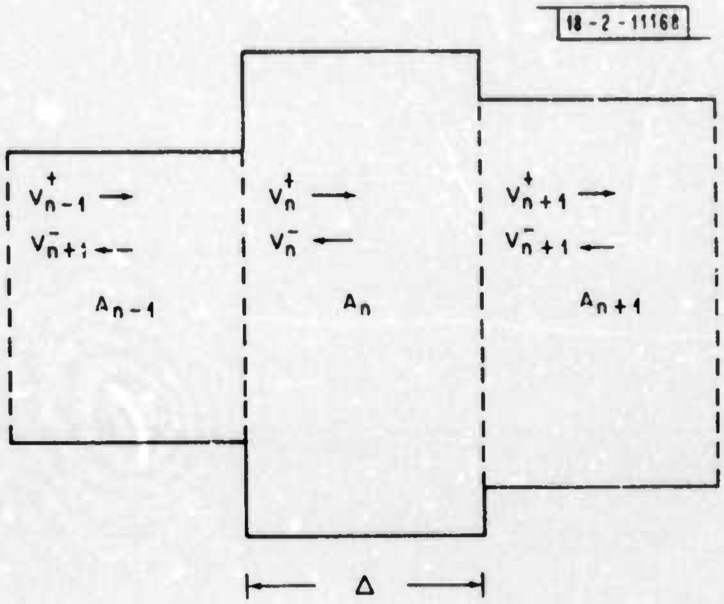

Fig. 1. Portion of a nonuniform acoustic tube.

In matrix form these equations read,

$$\begin{bmatrix} V_{n+1}^{+}(z) \\ \\ V_{n+1}^{-}(z) \end{bmatrix} = \frac{z^{-\frac{1}{2}}}{1+r_n} \begin{bmatrix} 1 & -r_n z \\ \\ -r_n & z \end{bmatrix} \begin{bmatrix} V_n^{+}(z) \\ \\ V_n^{-}(z) \end{bmatrix} \qquad (35)$$

Equation (35) can be inverted easily with the result,

$$\begin{bmatrix} V_n^{+}(z) \\ \\ V_n^{-}(z) \end{bmatrix} = \frac{z^{\frac{1}{2}}}{1-r_n} \begin{bmatrix} 1 & r_n \\ \\ r_n z^{-1} & z^{-1} \end{bmatrix} \begin{bmatrix} V_{n+1}^{+}(z) \\ \\ V_{n+1}^{-}(z) \end{bmatrix} \qquad (36)$$

These equations can be conveniently normalized by introducing the quantities

$$\begin{aligned} U_n^{+}(z) &= \frac{z^{\frac{n}{2}}}{\prod_{i=1}^{n-1}(1-r_i)} V_n^{+} \\ U_n^{-}(z) &= \frac{z^{\frac{n}{2}}}{\prod_{i=1}^{n-1}(1-r_i)} V_n^{-} \end{aligned} \qquad (37)$$

in terms of which equation (36) becomes:

$$\begin{bmatrix} U_n^{+}(z) \\ \\ U_n^{-}(z) \end{bmatrix} = \begin{bmatrix} 1 & r_n \\ \\ r_n z^{-1} & z^{-1} \end{bmatrix} \begin{bmatrix} U_{n+1}^{+}(z) \\ \\ U_{n+1}^{-}(z) \end{bmatrix} \qquad (38)$$

The quantities $U_n^{+}(z)$ and $U_n^{-}(z)$ can be interpreted as the forward and backward components of volume velocity in a fictitious acoustic tube which differs from the real tube only in that a gain factor $\prod_{i=1}^{n-1}(1-r_i)$ and an overall delay $z^{+\frac{n}{2}}$ have been removed.

Equation (38) can be used to derive a digital network whose response is the same as that of the acoustic tube. To accomplish this, equation (38) is first rewritten in the form:

$$U_{n+1}^{+}(z) = U_n^{+}(z) - r_n U_{n+1}^{-}(z)$$

$$U_n^{-}(z) = z^{-1}\left[r_n U_{n+1}^{+}(z) + U_{n+1}^{-}(z)\right] \quad (39)$$

The digital network that is generated by equation (39) is shown in Figure 2. This network as drawn is incomplete because no termination has been specified thus making it impossible to compute the sequence of backward going waves. As an example of a termination (one that will play a role in the sequel) assume the end of the tube is connected to a tube of infinite cross-section and of infinite length i.e., free space filled with air. This means that the final reflection coefficient is -1 and that there is no backward wave at the output. The network for this arrangement is shown in Figure 3 with the inputs to the network being the output of an N-section acoustic tube.

The next order of business is to compute the transfer function of an N-section acoustic tube. This will be done for the tube termination depicted in Figure 3 which implies that $U_{out}(z) = U_N^{+}(z)$. Since equation (38) enables one to recursively compute the z-transforms of the forward and backward waves in the n^{th} section of the tube in terms of their counterparts in the $n+1^{st}$ section it is natural to assume a simple output z-transform and then compute the input z-transform $U_1^{+}(z)$ that produced this output. If $U_{out}(z) = 1$ is assumed, then it follows that $U_N^{+}(z) = 1$ and $U_N^{-}(z) = -z^{-1}$. Equation (38) is now employed N times to arrive at $U_1^{+}(z)$ and it follows that the tube's transfer function is

$$T(z) = \frac{U_{out}(z)}{U_1^{+}(z)} = \left[U_1^{+}(z) \right]^{-1} \quad (40)$$

The computation just described is related to the Levinson recursion in a very important way. To make this fact clear, the Levinson recursion must be rewritten by introducing the functions $G_n^{+}(z)$ and $G_n^{-}(z)$ defined by

$$G_n^{+}(z) = H_n(z)$$

$$G_n^{-}(z) = -z^{-(n+1)} H_n(z^{-1}) \quad (41)$$

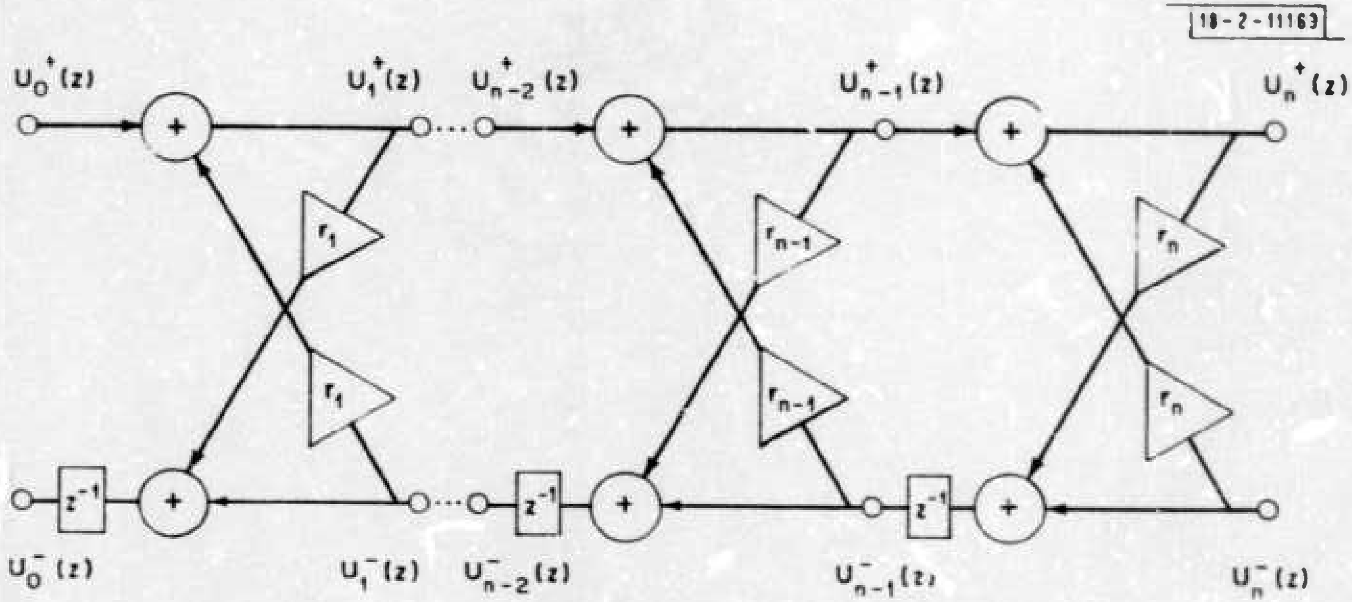

Fig. 2. Digital network for an acoustic tube.

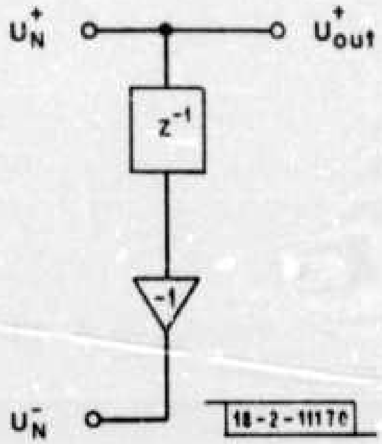

Fig. 3. Acoustic tube termination.

In terms of these functions, the Levinson recursion, equation (23), can be written as a set of two recursions as follows:

$$
\begin{aligned}
G^{+}_{n+1}(z) &= G^{+}_{n}(z) + K_n \, G^{-}_{n}(z) \\
G^{-}_{n+1}(z) &= z^{-1}\left[K_n \, G^{+}_{n}(z) + G^{-}_{n}(z) \right]
\end{aligned}
\qquad (42)
$$

or, in matrix form,

$$
\begin{bmatrix} G^{+}_{n+1}(z) \\ G^{-}_{n+1}(z) \end{bmatrix} = \begin{bmatrix} 1 & K_n \\ K_n z^{-1} & z^{-1} \end{bmatrix} \begin{bmatrix} G^{+}_{n}(z) \\ G^{-}_{n}(z) \end{bmatrix} \qquad (43)
$$

The initial condition for the recursion is now

$$
\begin{aligned}
G^{+}_{o}(z) &= 1 \\
G^{-}_{o}(z) &= -z^{-1}
\end{aligned}
\qquad (44)
$$

A comparison of equations (42) and (38) reveals that these two recursions are identical in form except that the indexing of the two are reversed, i.e., the acoustic tube indexing is from $n = N$ to $n = 1$ but the Levinson recursion indexes from $n = 1$ to $n = N$. Moreover, comparison of equation (44) and the initial conditions used for computing the acoustic tube's transfer function shows that these are also identical. What this all means is that an acoustic tube with reflection coefficients given by $r_{N-n} = K_{n-1}$ has a transfer function given by

$$
T(z) = \left[H_{N-1}(z) \right]^{-1}. \qquad (45)
$$

In other words, since the Levinson recursion yields the best estimate of the filter inverse to the filter that produced the original speech samples, the acoustic tube filter discussed above has a transfer function that is an estimate of the filter that originally produced the speech. Thus, this acoustic tube filter is a natural candidate for a filter to synthesize speech.

Atal (reference 8) has given a different derivation of the transfer function of a nonuniform acoustic tube. His derivation leads to the transfer function given by equation (45) however, his acoustic tube differs from the one derived above mainly in that the input and output terminals are interchanged. In other words, the reflection coefficient K_1 which appears at the output end of the acoustic tube derived above, appears at the input end of Atal's acoustic tube. Mathematically there does not seem to be any reason to choose one of these acoustic tubes over the other since they have identical transfer functions, however Wakita's tube seems more natural as a model of the vocal tract. This follows from the fact that Wakita's output termination is an infinite cross section tube which appears correct for modelling the interface between the lips and the outside world.

It has now been demonstrated how speech data can be used to derive a set of filter coefficients $a_k^{(p)}$ and a set of reflection coefficients K_n. The former could be used in a direct-form realization of a speech synthesis filter whereas the latter could be used to synthesize an acoustic tube synthesis filter. Which of these realizations is better is still a topic for investigation. For the sake of completeness, this section will conclude by showing how an <u>arbitrary</u>, stable all-pole filter $\left[H_n(z)\right]^{-1}$, can be realized as an acoustic tube.

The basic tool for this demonstration is the so-called backward Levinson recursion which can be derived from the forward Levinson recursion, equation (23) as follows. Solving equation (23) for $H_n(z)$ yields the relation,

$$H_n(z) = H_{n+1}(z) + K_n z^{-(n+1)} H_n(z^{-1}) \tag{46}$$

Next set $z = z^{-1}$ in equation (23) and solve again for $K_n H_n(z)$ with the result:

$$K_n H_n(z) = z^{-(n+1)} \left[H_n(z^{-1}) - H_{n+1}(z^{-1}) \right] \tag{47}$$

The elimination of $H_n(z^{-1})$ between equations (46) and (47) leads to the desired result:

$$H_n(z) = \frac{1}{1 - K_n^2} \left[H_{n+1}(z) + K_n z^{-(n+1)} H_{n+1}(z^{-1}) \right] \tag{48}$$

Since the constant term in $H_n(z)$ is unity, it follows from equation (48) that

$$a_{n+1}^{(n+1)} = K_n = - z^{(n+1)} H_{n+1}(z) \Big|_{z=0} \tag{49}$$

Let $H_N(z)$ denote an arbitrary N^{th} order polynomial in z^{-1} with constant term equal to unity. Furthermore, assume that all the zeros of $H_N(z)$ lie strictly inside the unit circle so that $\left[H_N(z) \right]^{-1}$ is the transfer function of a stable, all pole filter. Since all the zeros of $H_N(z)$ are inside the unit circle and since the coefficient of z^{-N} in $H_N(z)$ is the product of all the zeros of $H_N(z)$, it follows that K_N as given by equation (49) satisfies $|K_N| < 1$.

Assume next, that the backward Levinson recursion, equation (48), has been implemented n times and that $|K_{N-n+1}| < 1$ and that the polynomial $H_{N-n+1}(z)$ has a constant term equal to unity and that all its zeros lie inside the unit circle. It now follows from an application of Rouche's theorem that $H_{N-n}(z)$ as given by equation (48) has all of its zeros inside the unit circle and, therefore, that $|K_{N-n}| < 1$. The details of this argument will not be given here because they are virtually identical to those given earlier when it was shown that the forward Levinson recursion leads to stable filters as long as the K_n's used satisfy $|K_n| < 1$. It now follows by induction that all the K_n's produced by the backward Levinson recursion equations (48) and (49) satisfy $|K_n| < 1$ as long as the starting polynomial $H_N(z)$ had all of its zeros inside the unit circle.

Since it is obvious that a forward Levinson recursion using the K_n's derived from a backward Levinson recursion will yield back the starting polynomial $H_N(z)$, it follows from the discussion earlier in this section that a properly terminated acoustic tube having these K_n's as reflection coefficients will have a transfer function given by $\left[H_N(z)\right]^{-1}$. It has thus been shown how an arbitrary, stable all-pole filter can be realized as an acoustic tube.

The Orthogonal Polynomial Approach

The theory that has been presented is complete in itself, however, it should be pointed out that the results that have been derived are often arrived at in the literature by a completely different path making use of the theory of polynomials orthogonal on the unit circle*. The details of this alternate approach will now be presented. The first part of this section will deal exclusively with the theory of these polynomials with the connection to the speech problem being made later.

* This section is based on references 3, 4 and 5.

A weighting function w(z) is defined to be any function that satisfies $w(z) \geq 0$ on the unit circle and in addition, satisfies

$$\int w(z)\, df \quad > \ 0 \tag{50}$$

A finite or infinite set of polynomials,

$$\varphi_m (z) = \sum_{k=0}^{n} a_{nk} z^k \quad , \qquad n = 0, 1, \ldots \tag{51}$$

is said to be orthogonal with respect to the weighting function w(z) on the unit circle if

$$\left.\begin{aligned} &\text{a)} \quad a_{nn} > 0 \quad , \quad n = 0, 1, \ldots \\ &\text{b)} \quad \int \varphi_n (z)\ \overline{\varphi_m (z)}\ w(z)\, df \ = \ \delta_{nm} \end{aligned}\right\} \tag{52}$$

In equation (52), the overbar denotes complex conjugation and δ_{nm} the Kroneker delta.

It will now be shown that, given any weighting function, there exists a set of polynomials satisfying conditions a) and b). The proof will proceed by induction by defining,

$$\varphi_o (z) \ = \ c_o^{-\frac{1}{2}} \tag{53}$$

where

$$c_o \ = \int w(z)\, dz \tag{54}$$

The set of polynomials consisting of $\varphi_o(z)$ alone obviously satisfies a) and b).

Assume now that a set of N polynomials satisfying a) and b) has been constructed and enlarge this set by one by defining

$$\varphi_N(z) \ = \ A \left[z^N \ - \ \sum_{k=0}^{N-1} a_k \varphi_k (z) \right] \tag{55}$$

where A and the a_k's are to be determined.

It follows that

$$\int \varphi_N(z) \overline{\varphi_\ell (z)}\, w(z)\, df$$

$$= A\left[\int z^N \overline{\varphi_\ell(z)}\, w(z)\, df - a_\ell\right]$$

$$\ell = 0, \ldots\ N-1 \tag{56}$$

It is now obvious from equation (56) that condition b) will be satisfied by defining

$$a_\ell = \int z^N \overline{\varphi_\ell (z)}\, w(z)\, df$$

$$A = \left[\int \left| z^N - \sum_{k=0}^{N-1} a_k \varphi_k (z) \right|^2 w(z)\, df\right]^{-\frac{1}{2}} \tag{57}$$

The last equation is meaningful only if the integral appearing in it doesn't vanish which is always the case because it is well known that the powers of z form a linearly independent set. Finally, if the positive square root is always taken in equation (57), it follows that condition a) is also satisfied by the enlarged set of polynomials. The proof of existence is complete.

Next it will be shown that a set of polynomials satisfying a) and b) is unique. Assume the contrary. Then there exist two different sets of polynomials $\varphi_n(z)$ and $\varphi'_n(z)$ both satisfying a) and b). Next, note that it follows from condition b) that z^n can be written as a linear combination of $\varphi_n(z)$, $\varphi_{n-1}(z) \ldots$ $\varphi_0(z)$. (This is obvious for n = 0 and follows by a simple induction for the other powers of z.) This fact in turn implies that

$$\int \varphi_n(z)\, \bar{z}^k\, w(z)\, df = 0$$

$$k = 0, 1, \ldots\ n-1 \tag{58}$$

Now, because there are two sets of polynomials satisfying a) and b), it follows that the polynomial

$$p(z) = \varphi_n(z) - \frac{k_n}{k'_n} \varphi'_n (z) = 0 \tag{59}$$

where k_n and k'_n denote the coefficient of z^n in $\varphi'_n(z)$ and $\varphi'_n(z)$ respectively, is of degree no higher than n-1.

From this fact and equation (58), it follows that

$$\int \left| p(z) \right|^2 w(z)\, df$$

$$= \int \left[\varphi_n(z) - \frac{k_n}{k'_n} \varphi'_n(z) \right] \overline{p(z)}\, w(z)\, df \qquad (60)$$

$$= 0$$

and, therefore, that $p(z) = 0$ which implies that

$$\varphi_n(z) = \frac{k_n}{k'_n} \varphi'_n(z) \qquad (61)$$

However, $k_n = k'_n$ because,

$$1 = \int \left| \varphi_n(z) \right|^2 w(z)\, df$$

$$= \int \left| \varphi'_n(z) \right|^2 w(z)\, df \qquad (62)$$

and the uniqueness of any set of polynomials satisfying a) and b) has been established.

It is now possible to establish a number of important properties of orthogonal polynomials. The first of these is the fact that all the zeros of a set of polynomials satisfying a) and b) lie inside the unit circle. To prove this fact, let z_0 be a zero of $\varphi_n(z)$; $\varphi_n(z_0) = 0$. The polynomial $\varphi_n(z)/(z - z_0)$ is then of degree n-1 and it follows from equation (58) that

$$\int \varphi_n(z) \overline{\left[\frac{\varphi_n(z)}{z - z_0} \right]} w(z)\, df = 0 \qquad (63)$$

Equation (63) can easily be rewritten in the form,

$$\int (z - z_0) \left| \frac{\varphi_n(z)}{z - z_0} \right|^2 w(z)\, df = 0 \qquad (64)$$

from which it follows that,

$$z_0 = \frac{\int z \left| \frac{\varphi_n(z)}{z - z_0} \right|^2 w(z)\, df}{\int \left| \frac{\varphi_n(z)}{z - z_0} \right|^2 w(z)\, df} \qquad (65)$$

Since $z = z \cdot 1$, a simple application of the Schwartz inequality to equation (65) now shows that $|z_0| < 1$ where the strong inequality follows from the fact that z is not proportional to unity on the unit circle. This proves the theorem.

The next fact to be established provides the link between the theory of orthogonal polynomials and the speech problem introduced earlier. The property of orthogonal polynomials that accomplishes this is embodied in the statement that $\varphi_n(z)$ minimizes the integral

$$\int |p_n(z)|^2 \, w(z)\, df \tag{66}$$

where the minimum is taken over all polynomials of the form $p_n(z) = z^n + a_{n-1} z^{n-1} + \ldots a_0$. The minimum itself is k_n^{-2} where k_n denotes the coefficient of z^n in $\varphi_n(z)$.

The proof of this statement can be established by first noting that since z^n can be written as a linear combination of $\varphi_n(z)$, $\varphi_{n-1}(z), \ldots \varphi_0(z)$, it follows that any $p_n(z)$ can be represented as

$$p_n(z) = \sum_{k=0}^{n} V_k \varphi_k(z) \tag{67}$$

where $V_n = k_n^{-1}$ in order to force the coefficient of z^n in $p_n(z)$ to be unity. Substitution of equation (67) in equation (66) yields

$$\begin{aligned} \int |p_n(z)|^2 \, w(z)\, df &= \sum_{k=0}^{n} |V_k|^2 \\ &\geq |V_n|^2 = k_n^{-2} \end{aligned} \tag{68}$$

However, the lower bound given in equation (68) can be achieved by setting $V_k = 0$, $k = 0, \ldots n-1$ and the proof of the minimization property of orthogonal polynomials follows.

The connection to the speech problem now follows by recalling that this problem boiled down to minimizing the mean-squared error given by equation (4). Using Parseval's theorem, this equation can be rewritten in the z-transform domain with the result,

$$e = \int |H_p(z)|^2 \, |S(z)|^2 \, df \tag{69}$$

where

$$H_p(z) = 1 - \sum_{k=1}^{p} a_k z^{-k} \tag{70}$$

Since $|z^p| = 1$ on the unit circle minimizing the integral in equation (69) is the same as minimizing the integral given by

$$\int |z^p H_p(z)|^2 \; |S(z)| \; df \tag{71}$$

But $z^p H_p(z)$ is a p^{th} order polynomial with lead coefficient unity and it follows from the above minimization property of orthogonal polynomials that the minimum of (70) is given by k_p^{-2} and is achieved when

$$z^p H_p(z) = k_p^{-1} \varphi_p(z) \tag{72}$$

Here, $\varphi_p(z)$ denotes the p^{th} orthogonal polynomial with respect to the weighting function given by

$$w(z) = |S(z)|^2 \tag{73}$$

The above argument has transformed the speech problem under consideration from one of minimizing a certain integral to one of finding the p^{th} order orthogonal polynomial with respect to the weighting function $|S(z)|^2$. There exist explicit expressions for the polynomials orthogonal with respect to an arbitrary weighting function, however, their evaluation requires the computation of large determinants. A computationally more attractive approach to the evaluation of the coefficients of $\varphi_p(z)$ is available, however, because of the existence of a recursion formula for the orthogonal polynomials. The existence of such a recursion formula should come as no surprise; in fact, from the discussion in the previous section, it should be obvious that the desired recursion must be equivalent to the Levinson recursion. To derive this new version of the recursion, substitute equation (72) into the Levinson recursion, equation (23) with the result

$$k_{n+1}^{-1} \varphi_{n+1}(z) = k_n^{-1} z \varphi_n(z) - K_n k_n^{-1} z^n \varphi_n(z^{-1}) \tag{74}$$

Next the fact that k_n^{-2} is the mean squared error at the n^{th} stage coupled with equation (30) yields the final recursion formula

$$\varphi_{n+1}(z) = (1 - K_n^2)^{-\frac{1}{2}} \left[z\varphi_n(z) - K_n z^n \varphi_n(z^{-1}) \right] \qquad (75)$$

The K_n's appearing in equation (75) are still given by equation (21) where now

$$R_n = \int z^{-n} w(z)\, df \ . \qquad (76)$$

Conclusion

The basic mathematics relating to the linear predictive filtering approach to speech analysis/synthesis has now been presented. The analysis began by postulating that speech is produced by exciting an all-pole filter with either a uniform impulse train or white noise. A minimum mean-squared error technique for estimating the parameters of an all-pulse filter from a segment of speech data was then introduced and an explicit expression for this filter in terms of the speech data was derived.

Next, a numerically attractive recursive technique for computing this filter was derived and it was shown that this filter must always be stable. This filter can be realized in a variety of ways such as direct form, cascade form, and in addition, it was demonstrated that it also can be realized as a non-uniform acoustic tube. The reflection coefficients defining this tube are generated as a matter of course when computing the filter by means of the recursive technique just mentioned.

REFERENCES

1. H. Wakita, "Estimation of the Vocal Track Shape by Optimal Inverse Filtering and Acoustic/Articulatory Conversion Methods," Speech Communications Research Laboratory, Monograph No. 9, Santa Barbara, California.

2. J.I. Makhoul, Jared J. Wolf, "Linear Prediction and the Spectral Analysis of Speech," Bolt, Beranek and Newman Report No. 2304, Cambridge, Mass.

3. L.Ya. Geronimus, "Orthogonal Polynomials," Consultants Bureau, New York, 1961.

4. G. Szego, "Orthogonal Polynomials," American Mathematical Society Colloquium publication, Vol. XXIII, New York, 1959.

5. U. Grenander, G. Szego, "Toeplitz Forms and Their Applications," University of California Press, Berkeley and Los Angeles, 1958.

6. J.D. Markel, A.H. Gray, Jr., "Autocorrelation Equations as Applied to Speech Analysis," IEEE Transactions on Audio and Electroacoustics, Vol. AU-21, No. 2, April 1973.

7. N. Wiener, "Extrapolation, Interpolation and Smoothing of Stationary Time Series," The Technology Press and J. Wiley and Sons, New York, 1957, Appendix B.

8. B.S. Atal, S.L. Hanauer, "Speech Analysis by Linear Prediction of the Speech Wave," The Journal of the Acoustical Society of America, Vol. 50, No. 2 (Part 2), 1971.

9. F. Itakura, S. Saito, "Digital Filtering Techniques for Speech Analysis and Synthesis," Seventh International Congress on Acoustics, Budapest, 1971.

10. E.C. Titchmarch, "The Theory of Functions, Second Edition," Oxford University Press, 1939.